Development of a prediction model for infant hospitalization and death using clinical features assessed by community health workers during routine postnatal home visits in Dhaka, Bangladesh

Short title: Infant hospitalization and death prediction modeling

Alastair Fung[1,2,3,4], Marimuthu Sappani[5], Cole Heasley[2], Chun-Yuan Chen[2], Shaun K. Morris[1,2,3,4], Peter J. Gill[1,3,4], Diego G. Bassani[2,3,4], Davidson H. Hamer[6,7], Prakesh S. Shah[1,3,8], S. M. Abdul Gaffar[9], Sultana Yeasmin[9], Shafiqul A. Sarker[9], Shamima Sultana[9], Joseph Beyene*[3,5], Daniel E. Roth*[1,2,3,4]

*Contributed equally as co-senior authors

[1]Department of Paediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

[2]Centre for Global Child Health, The Hospital for Sick Children, Toronto, Canada.

[3]Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

[4]Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Canada.

[5]Department of Health Research Methods, Evidence and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Canada.

[6]Department of Global Health, Boston University School of Public Health, Boston, USA.

[7]Section of Infectious Diseases, Boston University Chobanian and Avedisian School of Medicine, Boston, USA.

[8]Department of Pediatrics, Mount Sinai Hospital and University of Toronto, Toronto, Canada.

[9]International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh.

ORCID

Alastair Fung: https://orcid.org/0000-0001-5615-4947

Marimuthu Sappani: https://orcid.org/0000-0002-8473-4241

Cole Heasley: https://orcid.org/0000-0002-1780-9856

Chun-Yuan Chen: https://orcid.org/0000-0002-6244-8455

Shaun K. Morris: https://orcid.org/0000-0001-9809-0822

Peter J. Gill: https://orcid.org/0000-0002-6253-1312

Diego G. Bassani: https://orcid.org/0000-0001-6704-3820

Davidson H. Hamer: https://orcid.org/0000-0002-4700-1495

Prakesh S. Shah: https://orcid.org/0000-0002-9920-0488

S. M. Abdul Gaffar: https://orcid.org/0000-0001-5625-1966

Sultana Yeasmin: https://orcid.org/0009-0004-8062-9868

Shafiqul A. Sarker: https://orcid.org/0000-0002-5566-2322

Shamima Sultana: https://orcid.org/0009-0000-7043-1621

Joseph Beyene: https://orcid.org/0000-0002-0770-4444

Daniel E. Roth: https://orcid.org/0000-0001-7742-0925

Correspondence: Alastair Fung, Division of Paediatric Medicine, The Hospital for Sick Children, 555 University Ave, Toronto, Ontario, M5G 1X8, Canada. Phone: 416-813-7500. Fax: 416-813-8418. Email: alastair.fung@sickkids.ca

Equal contribution: Joseph Beyene and Daniel Roth contributed equally as co-senior authors.

Abstract

Objectives

Community health worker (CHW) identification of life-threatening illnesses among young infants (<2 months) during home visits and referral to hospital are critical to reducing infant morbidity and mortality in low-resource settings. We aimed to develop a prediction model for hospitalization and/or death among young infants in Dhaka, Bangladesh using clinical features assessed by CHWs during routine home visits.

Methods

This was a secondary analysis of data from generally healthy infants prospectively enrolled at birth and assessed by CHWs at 11 scheduled home visits from 3-60 days of age. Time-varying Cox regression with backward selection was used to identify clinical features associated with time to first hospitalization and/or death. Prediction models were developed and internally validated using 5-fold cross-validation. We evaluated model discrimination (C-statistic and time-varying area under the curve) and calibration (calibration plots). We also evaluated discrimination and calibration of a Cox model based on World Health Organization (WHO)-recommended eight danger signs to identify sick infants requiring referral during home visits.

Results

Among 1906 infants, 176 (9.2%) had an event (173 hospitalizations and 3 deaths). The best-performing model consisted of three baseline covariates (any perinatal/delivery complication, umbilical cord care, gestational age) and four clinical features (nasal congestion, cough, jaundice, skin rash). The best-performing model discrimination (C-statistic=0.71; 95% CI 0.68-0.75), and discrimination of the best-performing model's four clinical features added to WHO

danger signs (C-statistic=0.70; 95% CI 0.67-0.74), were slightly higher than that of WHO danger signs alone (C-statistic=0.56; 95% CI 0.53-0.60), but calibration was similar.

Conclusions

A prediction model for hospitalization and/or death using baseline covariates and clinical features assessed during home visits may support identification of infants in need of facility-level care. Adding four clinical features to the WHO danger signs algorithm may improve its predictive performance by capturing a broader spectrum of severe infant illnesses requiring hospitalization.

Keywords: Prediction model, clinical signs, young infants, hospitalization, mortality, community health worker

Conflict of interest statement: The authors declare no conflicts of interest.

Funding statement: The SEPSiS observational cohort study was supported by the Bill and Melinda Gates Foundation (BMGF) grant number INV-007389 to The Hospital for Sick Children and grant number GR-02268 to The International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). BMGF had an advisory role in the overall study concept and design; however, the BMGF had no role in data collection, analysis, interpretation of data, writing of the article, and the decision to submit the article for publication. For the present study, AF was supported by the Canadian Institutes of Health Research (Funding Reference Number: FBD-181380) and the SickKids Clinician-Scientist Training Program (Project ID: 6030300488).

Acknowledgements: We thank Aline Freitas for contributing some of the statistical code that was used in or adapted for this analysis. We thank Safwan Ahmed and Vishva Shah for their assistance in reviewing references and verifying the accuracy and appropriateness of citations. We also thank all SEPSiS co-investigators and study personnel who contributed to the SEPSiS observational cohort study.

Word Count:

Abstract: 300 words

Manuscript: 3,498 words

Background

In 2023, 2.3 million infants worldwide died in the first month of age, with the highest burden in sub-Saharan Africa and central/southern Asia [1]. In low- and middle-income countries (LMICs), home-based postnatal care interventions, including community health worker (CHW) home visit assessments and referrals of sick newborns, have been shown to reduce neonatal mortality [2,3]. Infections are a leading cause of neonatal mortality, with sepsis accounting for 20% of neonatal deaths [4]. More than half of sepsis-related neonatal deaths in LMICs occur after the first week of age [5]. Therefore, home-based identification of life-threatening illnesses after the first week of age and referral to hospital are critical to reducing infant morbidity and mortality in LMICs.

The World Health Organization (WHO) recommends assessment of eight clinical signs ('WHO danger signs') by first-level health workers, including CHWs, during each postnatal home visit to promote timely identification of sick young infants (<2 months): not feeding well, history of convulsions, fast breathing (>60 breaths per minute), severe chest in-drawing, no spontaneous movement, fever (≥37.5^{o}C), low body temperature (<35.5^{o}C), and any jaundice in the first 24 hours of life or yellow palms and soles at any age [6,7]. Identification of any one sign indicates need for referral for further evaluation [7].

In a recent secondary analysis of a birth cohort in Bangladesh, India, and Pakistan (ANISA), each of seven signs of possible serious bacterial infection (pSBI; seven of the eight WHO danger signs, excluding jaundice) assessed by CHWs in young infants during scheduled home visits was significantly associated with mortality [8]. However, relying solely on WHO danger signs during routine visits presents potential challenges. First, WHO danger signs were originally derived among infants brought to health facilities due to concern [9]. These infants likely had higher

pretest probability of requiring hospital-based care than infants assessed during routine visits. Second, external validation of the WHO eight danger signs algorithm in a routine home visit setting [10] was limited to a small sample size (n=395) during the first week of age. Third, the algorithm had high specificity (95%) but lower sensitivity (69%) for physician-assessed severe illness requiring referral, raising concern that its use may miss cases [10]. Lastly, the danger signs were derived using single timepoint data rather than repeated (time-varying) measurements within each infant across home visits [9]. A prediction model derived using *visit-specific* time-varying predictors (e.g., occurrence of cough (yes/no) at each visit) accounts for changes in patient status over time, and may outperform single timepoint models (e.g., models using only measurements from the first visit) [11,12]. Beyond the infant's most recent visit, it is also possible to summarize repeated assessments of clinical signs at prior visits as *aggregative* time-varying predictors (e.g., cumulative number of visits with cough). Although WHO guidelines recommend multiple CHW home visits, the current algorithm considers the data from each visit in isolation, and does not utilize accumulated information. We hypothesized that a model using aggregative time-varying predictors may better reflect illness trajectory and improve predictive performance compared to a model based on the WHO danger signs, while acknowledging that aggregative time-varying predictors depend on repeated assessments which may affect feasibility. Therefore, we aimed to develop a prediction model for hospitalization and/or death among young infants in Dhaka, Bangladesh using time-varying clinical features assessed by CHWs during routine home visits.

Methods

Study design and data source

This was a secondary analysis of data from the Synbiotics for the Early Prevention of Severe

Infections in Infants (SEPSiS) observational cohort study (NCT04012190) [13]. From November 25, 2020 to February 18, 2022, mother-infant pairs were screened for eligibility at two government healthcare facilities in Dhaka: Maternal and Child Health Training Institute (Azimpur) and Mohammadpur Fertility Services and Training Centre. Infants born generally healthy were enrolled between day 0 (birth) and day 4 of age. This cohort (n=1939) has been described separately [14] and detailed inclusion/exclusion criteria are in **Panel S1**. Community health research workers (CHRWs) conducted infant assessments at up to 11 scheduled in-person home visits at 3- and 6-days post-enrolment, and on days 10, 14, 21, 28, 35, 42, 49, 56, and 60 postnatal age. If in-person visits were not feasible, assessments were attempted by telephone. Infants with at least one in-person CHRW visit after enrolment were included in this analysis. The observation period was from the first CHRW visit to day 67 of age, allowing predictors ascertained up to the 60-day visit to predict events up to day 67.

This secondary-use study was approved by The Hospital for Sick Children Research Ethics Board (REB #1000079158), and followed the TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis+Artificial Intelligence) guidelines [15]. Caregivers/public were not involved in study design, conduct, or dissemination.

Outcome

The outcome was time to first hospitalization and/or death (an event) following the first CHRW visit up to day 67 of age. If an infant was hospitalized and subsequently died, this counted as one event at time of hospitalization. If an infant had multiple hospitalizations, only the first hospitalization was included. Hospitalizations recommended by physicians but declined by caregivers were included. Hospitalizations prior to the first CHRW visit were excluded but these infants were included in the analysis following that visit. Two infants died before their first

CHRW in-person visit and were excluded. Hospitalizations for elective surgeries or trauma were excluded. Physicians recommending hospitalization were not study staff but they may have been informed by study physicians about CHRW findings.

Primary predictors

Primary candidate predictors were 45 clinical features (25 symptoms from caregiver history, 20 physical exam signs) assessed by CHRWs during scheduled home visits. The primary analysis only included clinical features from in-person visits. Using a standardized checklist, trained CHRWs assessed symptoms and signs occurring on the scheduled visit day, in the last 7 days, or since the last study visit (whichever was most recent). Each of the 45 clinical features was operationalized as visit-specific (i.e., based on data obtained at a single visit) and aggregative (i.e., based on accumulated data from prior visits) time-varying predictors (**Tables S1, S2, S3**). Candidate predictors included the WHO eight danger signs (which are, by definition, visit-specific). Per the SEPSiS protocol, CHRW ascertainment of any WHO danger sign mandated referral to a study physician.

Additional covariates

Additional covariates were included as predictors based on clinical judgment. Time-fixed additional covariates obtained at enrolment or the first scheduled visit included maternal age, maternal education, antenatal care, infant sex, gestational age at birth, preterm birth (<37 weeks), birth weight-for-gestational age Z-score, umbilical cord care, and any perinatal/delivery complication. Time-varying additional covariates included maternal postpartum substance use, exclusive breastfeeding status, and systemic antibiotic administration on the visit day, in the last 7 days or since the last visit (whichever was most recent). Details on the ascertainment and derivation of additional covariates can be found in **Panel S2**.

Sample size and missing data

Sample size was fixed based on available data from the SEPSiS study eligible for this secondary analysis. Missing values of clinical features were planned to be imputed, if necessary, using multiple imputation [16]; however, given minimal missingness of clinical features (1.2% to 1.4%), models were developed using complete case analysis.

Statistical analysis

Demographic characteristics were described using frequencies and percentages for categorical variables, and continuous variables were summarized by means and standard deviations, or medians and interquartile ranges.

Variables were selected for inclusion in prediction models based on clinical and statistical factors. Unadjusted time-varying Cox regression analyses were performed to estimate hazard ratios (HRs) with 95% confidence intervals (CI) and p-values. Prevalence of each feature was also considered during predictor selection: in the primary analysis, clinical features with prevalence ≥1% were used; the operationalization of each feature with the lowest p-value <0.2 in unadjusted analyses [17] was selected for multivariable analysis. Multicollinearity was assessed using the variance inflation factor (VIF) and variables with the highest VIFs ≥5 were sequentially removed until all VIFs were <5 [18].

Using predictors selected based on unadjusted analyses and multicollinearity, backward selection with a threshold p-value of <0.2 was used in a time-varying multivariable Cox regression analysis to derive a prediction model [17]. After backward selection, additional covariates and clinical features with p-value <0.2 were re-included in the model if this improved the C-statistic,

or removed if this did not change the C-statistic, to achieve the most accurate and parsimonious model. Internal validation was done using 5-fold cross validation [17].

Model performance was evaluated using discrimination (C-statistic and time-dependent area under the receiver operating characteristic curve (AUC) with 95% CIs generated by bootstrapping) and calibration (visual assessment of calibration plots) [19].

Additional analyses included: 1) evaluating discrimination and calibration of a time-varying Cox model based on WHO eight danger signs and a single predictor denoting at least one danger sign; 2) adding a single predictor denoting at least one WHO danger sign to the best-performing model in the primary analysis; 3) using only additional covariates and clinical features from caregiver history (excluding physical exam signs); 4) removing hospitalizations with a primary admission diagnosis of jaundice; 5) using a prevalence threshold of ≥0.1% instead of ≥1% for predictor inclusion in multivariable analysis; and 6) after backward selection, not re-including or removing additional covariates with p-value <0.2 from models if this improved discrimination. The rationale for the analysis excluding jaundice hospitalizations is that we anticipated a high proportion of hospitalizations for jaundice and a strong association between the jaundice predictor and these hospitalizations. We sought to evaluate whether model performance was retained for other clinically significant illnesses.

Sensitivity analyses included: 1) using a threshold p-value of <0.05 instead of <0.2 for backward selection; 2) excluding infants who received systemic antibiotic administration since the last study visit; 3) including events up to four weeks after the last CHRW visit; 4) excluding hospitalizations <48 hours duration; 5) imputing missingness with last observation carried forward; and 6) reducing the maximum number of scheduled visits.

Subgroup analyses included assessing discrimination of the best-performing model by 1) age group (0-28 days versus 28-60 days), 2) sex, and 3) maternal education level.

All analyses were done using R version 4.4.0 software [20].

Results

A total of 1906 infants had at least one CHRW home visit and were included (**Figure 1**). Of these, 176 had a hospitalization and/or death event (**Figure 1**). Admission diagnoses and causes of deaths are shown in **Tables S4 and S5**. CHRWs conducted 20,472 visits (82% in-person, 18% by telephone). The median (25^{th}, 75^{th}) number of visits (in-person and telephone) per infant was 11 (10, 11). Sixty-two percent of events occurred during the first four weeks of age and the median (25^{th}, 75^{th}) interval between an event and the most recent CHRW visit was 2 (1, 5) days (**Figures S1 and S2**). Of the additional covariates (**Table 1**), gestational age, any perinatal/delivery complication, umbilical cord care, and exclusive breastfeeding status and systemic antibiotic administration since the last visit met the p-value <0.2 threshold and were included in multivariable analyses.

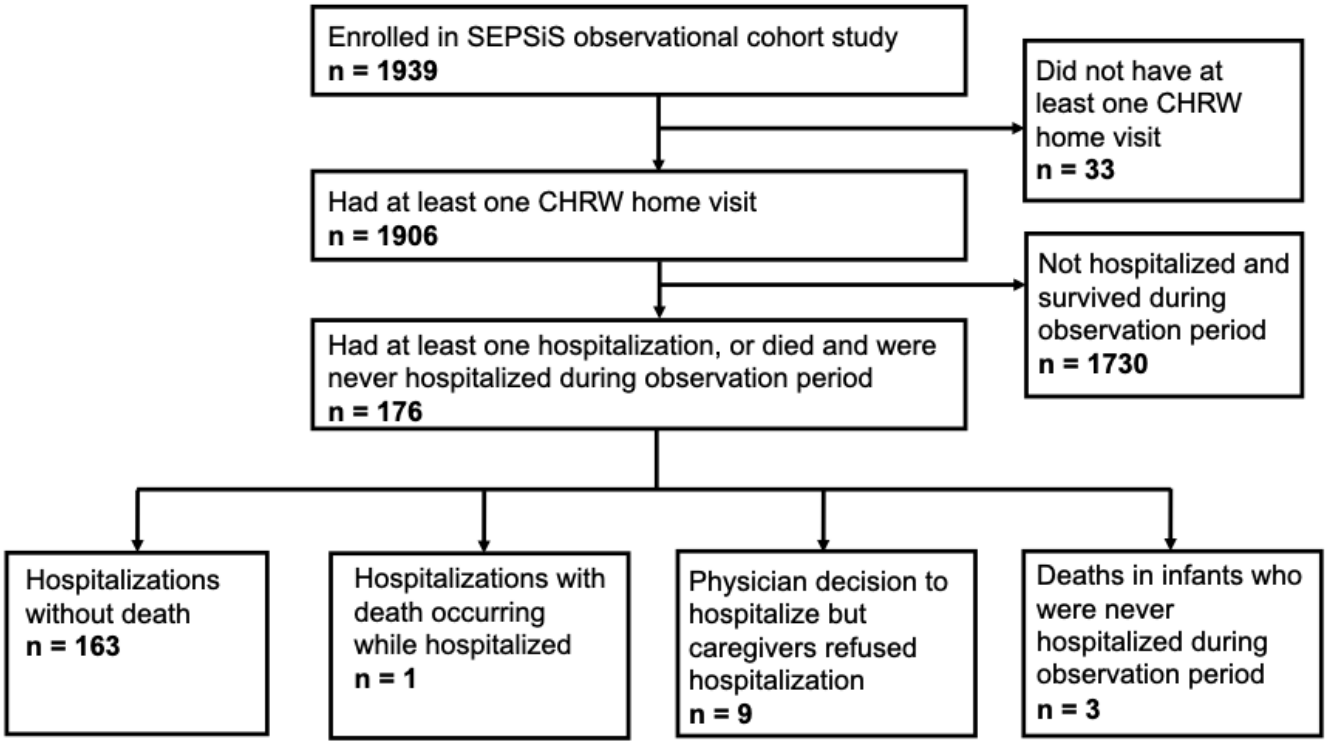


Figure 1. Study flow diagram

Table 1. Baseline characteristics and additional covariates and their associations with hospitalization and/or death in unadjusted analyses

Characteristic	Overall (n=1906)	No hospitalization or death (n=1730)	Hospitalized at least once or died without hospitalization (n=176)	Unadjusted HR (95% CI)	*P*-value
Maternal age (years), median (25th, 75th)[a]	24 (20, 27)	24 (20, 27)	23 (20, 27)	1.0 (0.97, 1.0)	0.99
Maternal education, n (%)[a]					
None up to complete primary school	533 (28)	475 (27)	58 (33)	Reference	
Secondary incomplete	612 (32)	561 (33)	51 (29)	0.75 (0.52, 1.1)	0.14
Secondary complete or higher	748 (40)	683 (40)	65 (37)	0.84 (0.59, 1.2)	0.35
Received antenatal care, n (%)[a]	1845 (97)	1677 (97)	168 (96)	1.3 (0.41, 4.0)	0.66
Female, n (%)	996 (52)	910 (53)	86 (49)	0.87 (0.64, 1.2)	0.34
Gestational age at birth (weeks), median (25th, 75th)[b]	39.1 (38.3, 40.1)	39.1 (38.3, 40.1)	38.9 (38.1, 39.9)	0.90 (0.83, 0.98)	0.021
Preterm birth (<37 weeks GA), n (%)	146 (7.7)	130 (7.5)	16 (9.1)	1.3 (0.76, 2.1)	0.36
Birth weight-for-gestational age Z-score, median (25th, 75th)	-0.81 (-1.5, -0.11)	-0.81 (-1.5, -0.11)	-0.86 (-1.5, -0.096)	0.96 (0.82, 1.1)	0.66
Any perinatal or delivery complication, n (%)[b,c]	107 (5.6)	93 (5.4)	14 (8.0)	1.6 (0.91, 2.7)	0.11
Umbilical cord care[b]					
None	276 (15)	262 (15)	14 (8.0)	Reference	
Antiseptic	1139 (60)	1023 (59)	116 (66)	2.1 (1.2, 3.6)	0.011
Antiseptic and other substance(s)[d]	374 (20)	340 (20)	34 (19)	1.7 (0.91, 3.2)	0.095
Other substance(s)[d]	117 (6.1)	105 (6.1)	12 (6.8)	2.0 (0.90, 4.2)	0.089
Exclusive breastfeeding status since the last study visit, n (%)[b,e]	1596 (84)	1455 (84)	141 (80)	0.72 (0.48, 1.1)	0.10

Any maternal postpartum substance use[f] since last study visit, n (%)[a,e]	132 (7.0)	124 (7.2)	8 (4.6)	0.59 (0.26, 1.4)	0.21
Systemic antibiotic administration since last study visit, n (%)[b],[e]	14 (0.73)	13 (0.75)	1 (0.57)	3.1 (2.0, 5.0)	<0.001

Missingness (0.68%-2.7%)

[a] Total number of mothers n=1893

[b] Met the p-value threshold of <0.2 and were selected as predictors in multivariable analyses

[c] Perinatal or delivery complication refers to any one or more of the following: maternal hemorrhage, premature rupture of membranes, prolonged rupture of membranes, foul smelling amniotic fluid, maternal fever, chorioamnionitis, maternal sepsis, fetal distress, sustained fetal bradycardia, sustained fetal tachycardia, or other complication.

[d] Other substance refers to any one or more of the following: lotion, desi ghee/butter, haldi (tumeric powder), mustard seed oil, coconut oil, powder/wheat flour, antimony (surmah), matti (soil/clay), unknown

[e] Time-varying variable. Proportions are based on data from first scheduled visit. Unadjusted HR and p-value account for data from all visits using time-varying Cox regression

[f] Postpartum substance use refers to use of any one or more of the following: smoked tobacco (cigarettes), Zarda (sweetened tobacco), tamak-pata (tobacco leaf), gul (dried and powdered tobacco), Pan (betel leaf), Supari (betel nut), Chuna (lime paste), unknown

Among the clinical features assessed by CHRWs, predictors selected using the p-value from unadjusted analyses and multicollinearity assessment for the primary analysis (lowest p-value <0.2 in unadjusted analysis, VIF<5, prevalence ≥1%) were all visit-specific time-varying operationalizations except for the cumulative sum of visits with red, oozing and/or swollen eyes up to the last visit, which was an aggregative time-varying predictor (**Table 2**). All other aggregative time-varying operationalizations had higher p-values, and hazard ratios attenuated to the null, compared with corresponding visit-specific variables and were therefore not carried forward to multivariable analysis (**Table 3**).

Table 2. Candidate predictors selected and their associations with hospitalization and/or death in unadjusted analyses

	Predictor	**Prevalence, n (%)[a]**	**Missing, n (%)[a]**	**Unadjusted HR (95% CI)**	***P*-value**
History	Stuffy nose	1142 (7.1)	192 (1.2)	4.9 (3.3-7.3)	<0.001
	Cough	691 (4.3)	192 (1.2)	5.6 (3.7-8.6)	<0.001
	Feels hot to touch or has fever within the past 7 days	508 (3.2)	0 (0.0)	2.6 (1.3-5.0)	0.0047
	Runny nose	227 (1.4)	193 (1.2)	5.5 (3.0-10)	<0.001
	Cumulative sum of visits with red, oozing and/or swollen eyes up to the last visit, median (range)[b]	0 (0, 8)	0 (0.0)	0.48 (0.21-1.1)	0.081
	Yellow discoloration of skin or eyes	199 (1.2)	193 (1.2)	12 (7.7-19)	<0.001
	Unusual skin rash or anything abnormal on skin	183 (1.1)	193 (1.2)	3.9 (1.9-7.9)	<0.001
Physical Exam	Axillary temperature (^{o}C), median (25^{th}, 75^{th})[c]	36.3 (36.1, 36.5)	203 (1.3)	1.9 (0.88-4.0)	0.11
	Cough	178 (1.1)	203 (1.3)	13 (8.1-22)	<0.001

Note: Using prevalence of predictors ≥1%, no variance inflation factors were >5.
[a] Denominator is total number of in-person visits (n=16,023)
[b] Value for first visit missing by default due to lagged variable and imputed with 0
[c] Mean of at least two and up to three temperature measurements during the same visit

Table 3. Comparison of associations of visit-specific and aggregative time-varying predictor operationalizations with hospitalization and/or death in unadjusted analyses (for predictors with prevalence ≥1%)

		VISIT-SPECIFIC		**AGGREGATIVE**					
		Occurrence at today's visit, in the past 7 days or since the last visit (whichever was most recent)		**Cumulative sum of visits with occurrence of clinical feature across all prior visits up to today's visit**		**Cumulative sum of visits with occurrence of clinical feature up to the last visit**		**Measurement at today's visit is at least 2 standard deviations above mean measurement across all prior visits**	
	Predictor	**Unadjusted HR (95% CI)**	***P*-value**	**Unadjusted HR (95% CI)**	***P*-value**	**Unadjusted HR (95% CI)**	***P*-value**	**Unadjusted HR (95% CI)**	***P*-value**
History	Stuffy nose	4.9 (3.3-7.3)	$1.3e^{-15}$	1.5 (1.3-1.7)	$1.8e^{-7}$	1.2 (0.98-1.5)	0.074	-	-
	Cough	5.6 (3.7-8.6)	$2.9e^{-15}$	1.7 (1.4-2.1)	$1.5e^{-7}$	1.3 (0.91-1.8)	0.16	-	-
	Feels hot to touch	2.6 (1.2-6.0)	0.021	1.3 (0.87-2.1)	0.19	0.95 (0.49-1.8)	0.87	-	-
	Runny nose	5.5 (3.0-10)	$5.1e^{-8}$	2.0 (1.3-2.9)	0.00064	1.2 (0.64-2.1)	0.60	-	-
	Red, oozing and/or swollen eyes	1.9 (0.79-4.8)	0.15	0.79, 0.53-1.2)	0.27	0.48 (0.21-1.1)	0.081	-	-
	Yellow discoloration of skin or eyes	12 (7.7-19)	$<2.0e^{-16}$	1.4 (1.3-1.6)	$1.8e^{-8}$	1.3 (0.88-1.9)	0.18	-	-
	Unusual skin rash or anything abnormal on skin	3.9 (1.9-7.9)	0.00023	1.6 (1.2-2.4)	0.016	1.3 (0.73-2.1)	0.42	-	-
Physical exam	Axillary temperature (^oC)	1.9 (0.88-4.0)	0.11	-	-	-	-	0.93 (0.50-1.7)	0.81
	Cough	13 (8.1-22)	$<2.0e^{-16}$	3.5 (2.4-4.9)	$1.6e^{-12}$	1.8 (0.96-3.5)	0.065	-	-

Note: Exact p-values are shown (rather than denoting as <0.001) to allow for visual comparison of p-values across operationalizations. As described in the Methods, aggregative time-varying predictors such as cumulative number of visits with cough, maximum temperature, and deviation of temperature above prior temperature values were operationalized within various time frames (e.g., past 2 weeks, past 4 weeks, or across all prior visits). The association of each operationalization for each clinical feature listed in **Table S3** with hospitalization and/or death was evaluated. For ease of interpretation, only certain operationalizations are shown in this table.

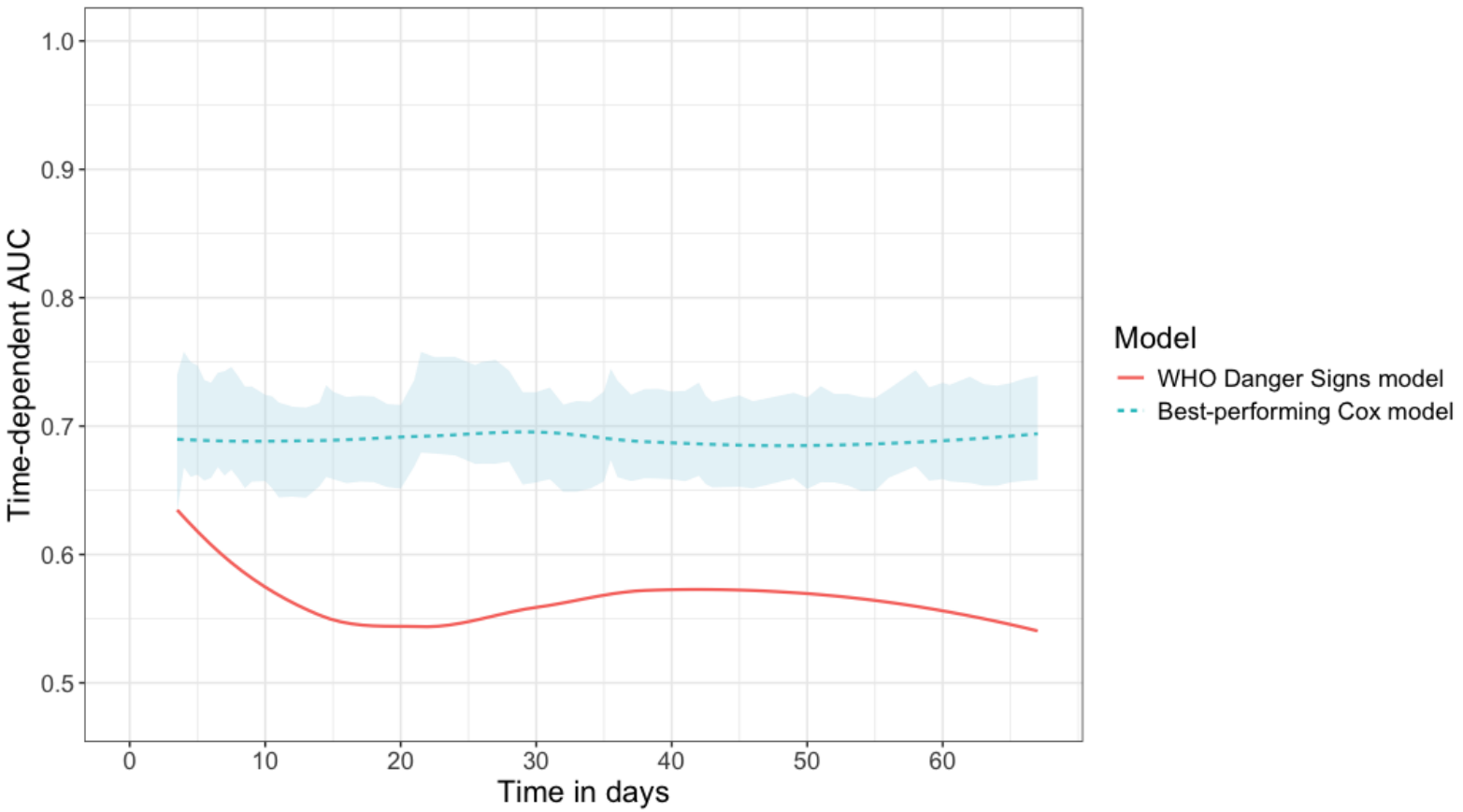
The best-performing multivariable Cox model included seven predictors and had a C-statistic of 0.71 (95% CI 0.68-0.75) and a similar cross-validated C-statistic (**Table 4**). The model had a consistent time-dependent AUC across the observation period (**Figure 2**).

Table 4. Discrimination of best-performing Cox model for hospitalization and/or death among young infants (using predictors with prevalence ≥1%)

	Predictor	**Prevalence, n (%)[a]**	**Adjusted HR (95% CI)**	***P*-value**	**C-statistic (95% CI)**	**Cross-validated C-statistic (95% CI)**
Additional covariates	Any perinatal or delivery complication	864 (5.4)	1.3 (0.71-2.2)	0.43	0.71 (0.68-0.75)	0.68 (0.58-0.78)
	Umbilical cord care					
	Antiseptic	9546 (60)	1.6 (0.92-2.8)	0.096		
	Antiseptic and other substance(s)	3241 (20)	1.5 (0.79-2.8)	0.23		
	Other substance(s)	941 (5.9)	1.6 (0.71-3.5)	0.26		
	Gestational age at birth (weeks), median (25th, 75th)[b]	39.1 (38.3, 40.1)	1.1 (1.0-1.2)	0.032		
History	Stuffy nose	1142 (7.1)	3.0 (2.0-4.6)	<0.001		
	Yellow discoloration of skin or eyes	199 (1.2)	11 (6.6-17)	<0.001		
	Unusual skin rash or anything abnormal on skin	183 (1.1)	2.3 (1.1-4.8)	0.024		
Physical exam	Cough	178 (1.1)	7.5 (4.4-13)	<0.001		

[a] Denominator is total number of in-person visits (n=16,023)
[b] Adjusted hazard ratio inverted so that lower gestational age corresponds to higher risk

Note: The cumulative sum of visits with red, oozing and/or swollen eyes up to the last visit prior to the current visit met the criteria to be selected as a candidate predictor and was included after backward selection (**Table 2**). Since this predictor was the only aggregative time-varying operationalization in the final set of candidate predictors, its removal would support the feasibility of the final best-performing multivariable time-varying Cox model. Removing this predictor changed the C-statistic of the final multivariable model from 0.72 (95% CI 0.69-0.76) to 0.71 (95% CI 0.68-0.75) and it was therefore removed to achieve a more parsimonious and feasible final multivariable model.



AUC: Area under the receiver operating characteristic curve; WHO: World Health Organization.

Note: Shaded area represents 95% confidence interval (CI). The 95% CI for WHO danger signs is not shown because the CI was very wide and not useful for illustrative purposes.

Figure 2. Time-dependent area under the receiver operating characteristic curve (AUC) for best-performing Cox model for hospitalization and/or death among young infants (using prevalence of predictors ≥1%) and Cox model based on World Health Organization danger signs

The WHO eight danger signs were infrequent (<0.1% of visits). A Cox model based on WHO danger signs or a single variable denoting at least one of these signs both had a C-statistic of 0.56 (0.54-0.60) with a similar cross-validated C-statistic (**Table 5**). The time-dependent AUC value of the WHO danger signs was higher during the first week of age and subsequently decreased (**Figure 2**).

Table 5. Discrimination of Cox model for hospitalization and/or death among young infants based on World Health Organization danger signs

	Predictor	**Prevalence, n (%)[a]**	**Adjusted HR (95% CI)**	***P*-value**	**C-statistic (95% CI)**	**Cross-validated C-statistic (95% CI)**
History	Abnormal movement (convulsions/fits)	9 (0.056)	15 (2.1-107)	0.0073	0.56 (0.54-0.60)	0.51 (0.47-0.55)
Physical exam	Not sucking effectively	17 (0.11)	14 (4.2-46)	<0.001		
	Fast breathing (RR >60 per minute)	19 (0.12)	2.3 (0.48-12)	0.30		
	Severe lower chest wall in-drawing	14 (0.087)	20 (5.0-81)	<0.001		
	Lethargy (no movement or movement only on stimulation)	4 (0.025)	48 (8.4-275)	<0.001		
	Fever (one temperature measurement >38^{o}C or two measurements ≥37.5^{o}C)	11 (0.069)	32 (9.9-102)	<0.001		
	Low body temperature (one temperature measurement <34.5^{o}C or two measurements <35.5^{o}C)	7 (0.047)	2.4 (0.13-43)	0.56		
	Yellow discoloration of hands and soles of feet	13 (0.081)	31 (12-78)	<0.001		
Physical exam	1 of any of the 8 danger signs	83 (0.52)	20 (12-33)	<0.001	0.56 (0.53-0.60)	0.54 (0.49-0.58)

[a] Denominator is total number of in-person visits (n=16,023)

The best-performing model with the addition of a variable denoting at least one WHO danger sign had a C-statistic of 0.73 (0.69-0.76) (**Table S6**). A model consisting of the four clinical features from the best-performing model plus a variable denoting at least one WHO danger sign had a C-statistic of 0.70 (0.67-0.74) (**Table 6**).

Table 6. Discrimination of best-performing Cox model for hospitalization and/or death among young infants limited to clinical features (with additional covariates removed) plus a variable denoting at least one WHO danger sign

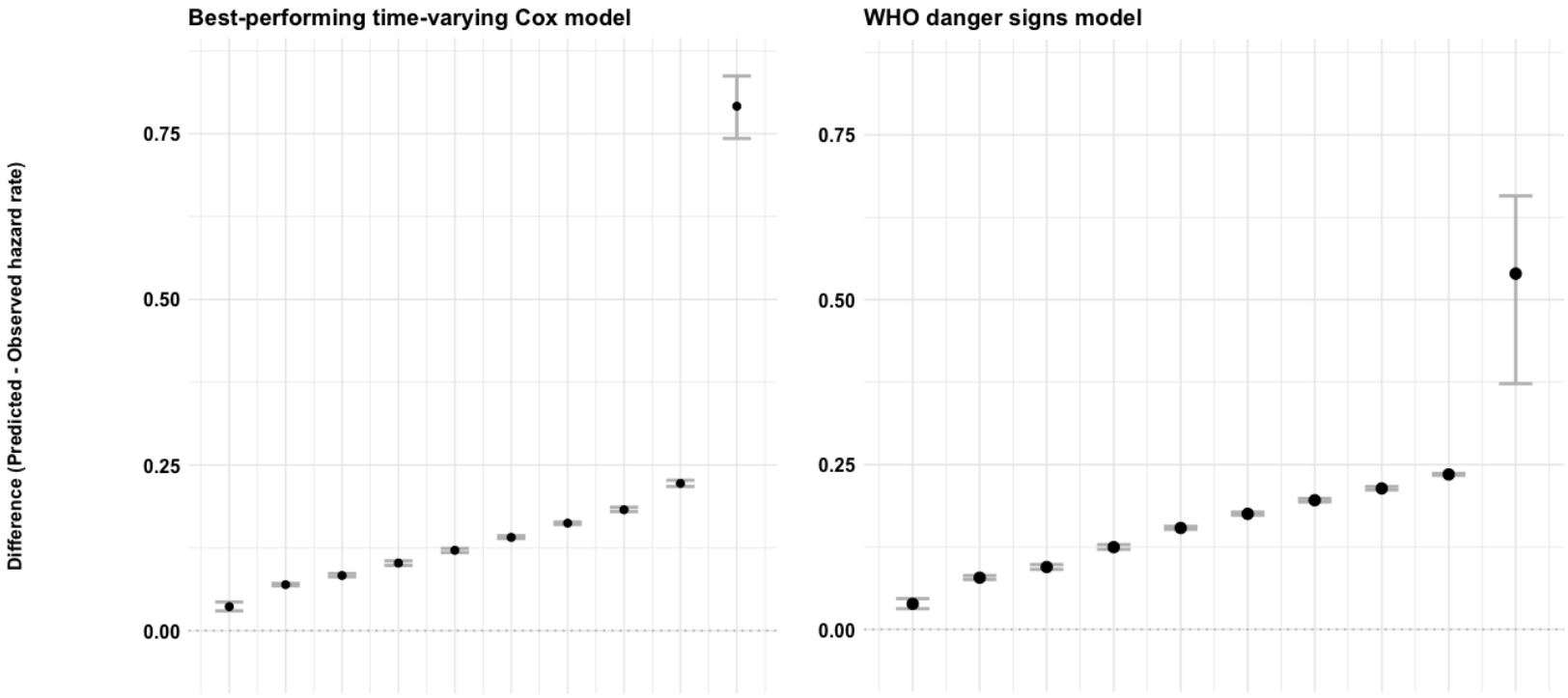
	Predictor	**Prevalence, n (%)[a]**	**Adjusted HR (95% CI)**	***P*-value**	**C-statistic (95% CI)**	**Cross-validated C-statistic (95% CI)**
History	Stuffy nose	1142 (7.1)	3.0 (2.0-4.5)	<0.001	0.70 (0.67-0.74)	0.68 (0.59-0.76)
	Yellow discoloration of skin or eyes	199 (1.2)	8.9 (5.5-14)	<0.001		
	Unusual skin rash or anything abnormal on skin	183 (1.1)	2.9 (1.4-5.9)	0.0044		
Physical exam	Cough	178 (1.1)	5.6 (3.1-10)	<0.001		
	1 of any of the 8 WHO danger signs	83 (0.52)	6.0 (3.4-11)	<0.001		

[a] Denominator is total number of in-person visits (n=16,023)

In an additional analysis limiting predictors to baseline additional covariates and caregiver-reported symptoms alone during in-person visits, the best-performing model discrimination was similar (**Table S7**). When the additional covariates in the best-performing model were removed, leaving only four clinical features, the C-statistic decreased slightly (**Table S8**).
Other additional, sensitivity, and subgroup analyses demonstrated no substantial changes in discrimination (**Tables S9, S10, S11, S12**).

The best-performing model (**Table 4**) was derived using data from up to 11 CHRW visits. As the maximum number of CHRW home visits per infant decreases, thereby increasing the average prediction window (i.e., number of days between an event and the most recent visit), the C-statistic decreases (**Figure S3**).

Across all deciles of predicted risks, both the best-performing model and WHO danger signs model overestimated predicted risks with gradually increasing differences between predicted and observed hazard rates and 95% CIs excluding 0 (**Figure 3**).



Decile	1	2	3	4	5	6	7	8	9	10
Mean predicted hazard rate	0.053	0.073	0.088	0.11	0.12	0.14	0.16	0.19	0.23	0.83
Mean observed hazard rate	0.017	0.0031	0.0051	0.0041	0.0033	0.0029	0.0014	0.0039	0.0027	0.035

1	2	3	4	5	6	7	8	9	10
0.066	0.084	0.10	0.13	0.16	0.18	0.20	0.22	0.24	0.55
0.027	0.0052	0.0076	0.0062	0.0040	0.0042	0.0032	0.0036	0.0021	0.015

Discrete risk groups were defined by deciles of the predicted values with the first decile having the lowest predicted values and the 10th decile having the highest predicted values. Calibration was assessed by risk group. For each decile risk group, the mean predicted value was compared with the observed value. The difference between the predicted and observed hazard rates was computed and then plotted for each corresponding decile. The gray bars indicate the 95% confidence intervals from 1000 bootstrapped data sets. For well-calibrated models, the difference between the predicted and observed hazard rates should be minimal (near 0) [19].

Figure 3. Calibration plots by decile of predicted hazard rates for the best-performing Cox model and the Cox model based on World Health Organization danger signs.

Discussion

A Cox model using time-fixed baseline covariates and visit-specific time-varying clinical features demonstrated higher discrimination and similar calibration compared to the conventional approach using WHO danger signs to predict hospitalization and/or death among young infants during CHW routine home visits in Dhaka, Bangladesh. Discrimination was maintained using four clinical features from the best-performing model (nasal congestion, jaundice, and skin rash on history, and cough on physical exam) plus a variable denoting at least one WHO danger sign. Aggregative time-varying operationalizations were less associated with hospitalization and/or death than visit-specific operationalizations, and were not retained in the best-performing model.

In a 2024 systematic review of the diagnostic accuracy of clinical sign algorithms to identify sepsis in young infants [21], the most accurate algorithm in the routine home visit setting was the WHO danger signs. None of the included studies developed time-varying models. Although WHO danger signs had poorer discrimination than the best-performing model in the present study, they remain important given their high specificity (~95%) for severe illness requiring referral [10] and their significant association with mortality [8]. However, in this urban Dhaka cohort of infants born generally healthy (excluding those on parenteral antibiotics at the time of eligibility screening), WHO danger signs were infrequent at routine home visits (<0.1% of visits). In contrast, in a secondary analysis of ANISA, which was a birth cohort and included rural areas, the prevalence of at least one sign of pSBI in young infants during CHW scheduled home visits was substantially higher at 2.7% [8]. In the present study, clinical features in the best-performing model, including nasal congestion, jaundice, skin rash, and cough, were much more frequently ascertained during routine visits than WHO danger signs. Forty-two percent of hospitalizations were attributed to pneumonia for which milder and more common symptoms

such as nasal congestion and cough may have higher predictive accuracy than WHO danger signs. Furthermore, discrimination improved when combining the best-performing model, or the four clinical features from the best-performing model, with a variable denoting any one WHO danger sign compared to the danger signs alone. Therefore, WHO danger signs remain essential but may be insufficient to accurately identify the full spectrum of illnesses requiring hospitalization in young infants in an urban setting. Adding the four clinical features identified in the best-performing model to the WHO danger signs algorithm may improve its predictive performance.

We hypothesized that aggregative time-varying predictors would have higher predictive accuracy than their corresponding visit-specific operationalizations. From a clinical standpoint, cumulative days or weeks of cough would be more concerning than cough measured once on the visit day. Although there were up to 11 scheduled home visits (median of 11 visits attended), this frequency of follow-up may not have been high enough to capture the rapid progression of severe illnesses requiring hospitalization in young infants. For example, if a pneumonia evolved over the course of days 38 to 42 of age, the infant may not have exhibited cough or other concerning symptoms and signs during the recent scheduled preceding visits on days 28 and 35, making the predictor ‘cumulative number of visits with cough’ uninformative. The onset and progression of severe illnesses such as sepsis would also be expected to be rapid (within hours to a few days), making it challenging for aggregative operationalizations of symptoms and signs across visits scheduled four to seven days apart to reliably detect sepsis.

A limitation of this study was that the observation period started at the first CHRW home visit (day 3 to 7 of age), thereby excluding the first 24 hours of age when neonatal mortality is highest [22]. However, infants in this setting remained at high risk of hospitalization and death from

days 3 to 67 of age, substantiating the importance of prediction models applicable to this period. Second, prevalence of symptoms and signs during routine home visits was relatively low. In the urban setting of Dhaka where health facilities are numerous and caregivers readily seek care for infant illness, CHWs are less likely to ascertain concerning features during routine home visits. In a separate analysis of severe infection incidence in this cohort, 85% of severe infection episodes were identified following caregiver self-referral compared to 15% identified following a scheduled CHRW home visit assessment and referral [14]. External validation of the model is needed, particularly in settings with more limited access to health facilities and where caregivers may not as readily seek care for infant illness (e.g., rural communities). Third, the baseline additional covariates in the best-performing model may not be readily available outside of a research setting. For example, gestational age at birth was calculated by an electronic data capture system by averaging the time since the first day of the last menstrual period in the mother's antenatal card (if available) and the estimated gestational age from the ultrasound report (if available). Nevertheless, a sensitivity analysis removing the baseline additional covariates did not demonstrate a substantial decrease in discrimination. Fourth, the protocol-defined referral system, mandating referral to a study physician if a CHRW ascertained one of the WHO danger signs during a home visit, may have increased the association of the danger signs with hospitalization, thereby potentially introducing a bias that may improve discrimination of the WHO danger signs model. However, despite this potential protocol-induced bias, the best-performing model did not include any of the WHO danger signs and showed higher discrimination than the WHO danger signs model. Fifth, although physicians assessing the outcome were not study staff, there was no protocol-instituted blinding to CHRW assessments, and their recommendations for hospitalization could have been influenced by CHRW findings.

Implications for clinical use and future research

With a C-statistic of 0.71, the best-performing model may lack sufficient discrimination in its current form to reliably guide CHW referral decisions. Furthermore, the best-performing model was derived using up to 11 CHRW home visits and requires baseline data which may be considered too resource-intensive in many settings. A trend of decreasing discrimination was observed as the maximum number of home visits decreased. However, when the maximum number of visits was reduced to four, as recommended by current WHO guidelines [7], model discrimination was similar. Therefore, although using up to 11 visits increases model discrimination, this high number of visits may not be essential. When implementing this model, one must consider the trade-off between the labour and costs associated with the number of CHW visits and model discrimination.

If in-person CHW visits are feasible, a combination of the best-performing model plus a variable denoting any one WHO danger sign demonstrated the highest discrimination. Considering the potential challenges with ascertaining baseline covariates such as gestational age in LMIC settings, removing the additional covariates from the best-performing model and simplifying it to its four clinical features plus a variable denoting any one WHO danger sign showed similar discrimination. However, before this simplified model plus WHO danger signs can be implemented into practice, it requires external validation. Future studies should externally validate the simplified model plus WHO danger signs in both urban and rural LMIC settings. Lastly, backward selection using a threshold p-value is commonly used for predictor selection [23], but other methods exist [24,25,26] and may be considered in future studies.

Conclusions

A prediction model for infant hospitalization and/or death in Dhaka, Bangladesh using baseline covariates and visit-specific time-varying clinical features assessed by CHWs during routine home visits may support identification of sick young infants in LMICs. Aggregative time-varying clinical features do not improve prediction of hospitalization and/or death compared to visit-specific clinical features. Adding four clinical features – nasal congestion, jaundice, skin rash, and cough – to the WHO danger signs algorithm may improve its predictive performance by capturing a broader spectrum of severe illnesses requiring hospitalization in young infants.

Supporting information: Supplementary Tables and Figures can be found in the **Supplemental File**.

Study protocol: The protocol for this study is available upon request.

Ethics statement: This secondary analysis study of the SEPSiS Observational Cohort Study was approved by The Hospital for Sick Children Research Ethics Board (REB #1000079158). The SEPSiS Observational Cohort Study was approved by The Hospital for Sick Children Research Ethics Board (REB #1000063899) and the ethical review committees at the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) (PR-19045) and Bangladesh Shishu Hospital and Institute (formally known as Dhaka Shishu Hospital), the ethical governing body for the Child Health Research Foundation (CHRF) (BICH-ERC-20/02/2019). Informed written consent was obtained from a parent or legal guardian before participant enrolment.

Data availability: De-identified datasets and code files used in the analyses of this study are publicly available at the Borealis online data repository at https://doi.org/10.5683/SP3/AMBJCL under Custom Dataset Terms. Standard operating procedures (SOPs), including those related to SEPSiS observational cohort study enrolment, consent and data collection procedures, are available at https://doi.org/10.5683/SP3/WKDQYY.

Authorship contributions: AF, DGB, SKM, PJG, DHH, PSS, SMAG, SY, SAS, SS, JB, and DER conceptualized the study and developed the methodology. AF, MS, CH, and C-YC managed the data and conducted the data analysis. JB and DER supervised data analysis. AF wrote the initial draft of the manuscript. All authors provided feedback on data analysis and the manuscript. All authors read and approved the final manuscript.

References

1. UNICEF and partners in the United Nations Inter-Agency Group for Child Mortality Estimation. Levels and Trends in Child Mortality Report 2024: Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation: United Nations Children's Fund; 2025 [Available from: https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2024/.

2. Tiruneh GT, Shiferaw CB, Worku A. Effectiveness and cost-effectiveness of home-based postpartum care on neonatal mortality and exclusive breastfeeding practice in low-and-middle-income countries: a systematic review and meta-analysis. BMC Pregnancy Childbirth. 2019;19(1):507.

3. Lassi ZS, Kedzior SG, Bhutta ZA. Community-based maternal and newborn educational care packages for improving neonatal health and survival in low- and middle-income countries. Cochrane Database Syst Rev. 2019;2019(11).

4. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet. 2020;395(10219):200-11.

5. Sankar MJ, Natarajan CK, Das RR, Agarwal R, Chandrasekaran A, Paul VK. When do newborns die? A systematic review of timing of overall and cause-specific neonatal deaths in developing countries. J Perinatol. 2016;36 Suppl 1(Suppl 1):S1-s11.

6. World Health Organization. Integrated Management of Childhood Illness: management of the sick young infant aged up to 2 months Geneva: World Health Organization; 2019 [Available from: https://www.who.int/maternal_child_adolescent/documents/management-sick-young-infant-0-2-months/en/.

7. WHO recommendations on maternal and newborn care for a positive postnatal experience Geneva: World Health Organization; 2022 [Available from: https://www.who.int/publications/i/item/9789240045989.

8. Darmstadt GL, Ahmed S, Islam MS, Abdalla S, El Arifeen S, Arvay ML, et al. Association of clinical signs of possible serious bacterial infections identified by community health workers with mortality of young infants in South Asia: a prospective, observational cohort study. EClinicalMedicine. 2025;80:103070.

9. Young Infants Clinical Signs Study Group. Clinical signs that predict severe illness in children under age 2 months: a multicentre study. Lancet. 2008;371(9607):135-42.

10. Darmstadt GL, Baqui AH, Choi Y, Bari S, Rahman SM, Mannan I, et al. Validation of a clinical algorithm to identify neonates with severe illness during routine household visits in rural Bangladesh. Arch Dis Child. 2011;96(12):1140-6.

11. Fung A, Loutet M, Roth DE, Wong E, Gill PJ, Morris SK, et al. Clinical prediction models in children that use repeated measurements with time-varying covariates: a scoping review. Acad Pediatr. 2024;24(5):728-40.

12. Wen B, Brals D, Bourdon C, Erdman L, Ngari M, Chimwezi E, et al. Predicting the risk of mortality during hospitalization in sick severely malnourished children using daily evaluation of key clinical warning signs. BMC Med. 2021;19(1):222.

13. SEPSIS Observational Cohort Study in Young Infants in Bangladesh [Available from: https://ClinicalTrials.gov/show/NCT04012190.

14. Fung A, Heasley C, Pell LG, Bassani DG, Shah PS, Morris SK, et al. Severe infection incidence among young infants in Dhaka, Bangladesh: an observational cohort study. BMJ Public Health. 2025;3(2):e002383.

15. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. Bmj. 2024;385:e078378.

16. Austin PC, White IR, Lee DS, van Buuren S. Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Can J Cardiol. 2021;37(9):1322-31.

17. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics. 2nd ed. Boston, MA: Springer; 2012.

18. Lavery MR, Acharya P, Sivo SA, Xu L. Number of predictors and multicollinearity: What are their effects on error and bias in regression? Communications in Statistics - Simulation and Computation. 2019;48(1):27-38.

19. Steyerberg EW. Chapter 15.3: Calibration. Clinical Prediction Models: A Practical Approach to Development, Validating, and Updating. Statistics for Biology and Health. 2nd ed. Switzerland: Springer; 2019.

20. R Core Team. R: A Language and Environment for Statistical Computing Vienna, Austria.: R Foundation for Statistical Computing; 2025 [Available from: https://www.R-project.org/.

21. Fung A, Shafiq Y, Driker S, Rees CA, Mediratta RP, Rosenberg R, et al. Diagnostic Accuracy of Clinical Sign Algorithms to Identify Sepsis in Young Infants Aged 0 to 59 Days: A Systematic Review and Meta-analysis. Pediatrics. 2024;154(Suppl 1).

22. Dol J, Hughes B, Bonet M, Dorey R, Dorling J, Grant A, et al. Timing of neonatal mortality and severe morbidity during the postnatal period: a systematic review. JBI Evid Synth. 2023;21(1):98-199.

23. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683-90.

24. Steyerberg EW. Chapter 11: Selection of Main Effects. Clinical Prediction Models: A Practical Approach to Development, Validating, and Updating. Statistics for Biology and Health. 2nd ed. Switzerland: Springer; 2019. p. 207-25.

25. Austin PC, Tu JV. Bootstrap Methods for Developing Predictive Models. The American Statistician. 2004;58(2):131-7.

26. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological). 2018;58(1):267-88.